**CSCE 623: Machine Learning**

**Spring 2021**

**PROJECT FIRST DRAFT PAPER GRADING WORKSHEET**

Due Tue, 18 May at 2359

Submit via Canvas

**(**This draft is worth 10 points toward your final grade: 8 for completeness and 2 for quality**)**

|  |  |
| --- | --- |
| Course points earned | 9.0 |

**Dartt\_Rain\_F\_Project\_FirstDraft\_Grading\_Worksheet.docx**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Criteria | Sub Area | Avail points | Student performance | Student Score |
| Completeness | INTRODUCTION  - motivate the problem  - clear research question/hypothesis/objective | 2 | Achieved | 2 |
| Completeness | RELATED WORK  - minimum = course text | 1 | Achieved | 1 |
| Completeness | APPROACH / METHODOLOGY  - Describe/explain/explore data and preprocessing steps (What is the nature of the data (or experiments that generate the data) you are working with?)  - models defined  - hyper-parameters & validation scheme  - performance measures defined  - comparison to what baseline?  - Defined analysis techniques | 3 | Partial: see below | 2.2 |
| Completeness | (EXPECTED) RESULTS | 1 | OK but see below | 1 |
| Completeness | REFERENCES  - minimum = course text  - citations in main text  - ref section | 1 | Achieved | 1 |
| Quality | Length (1000-3000 words) Spelling/Grammar | 1 | Good | 1 |
| Quality | Logic / Flow | 1 | Partial – lots of deviation from good tech writing behaviors… see below | 0.8 |
| TOTAL |  | 10 |  | 9.0 |

**Other Comments:**

“Currently, AFPC relies on human technicians who are not trained or familiar with field-specific jargon that often appears in these resumes, so misclassification is too common, resulting in time wasted on false positives and accusations of unfairness (and worse) from false negatives.” -> do you have any numbers to go with this (or estimates like 20% false positive rate)? it would be grate to have something specific so that you can determine whether you are performing better than the alternative and could state specifically how your model is doing.

DST/Workflow… don’t get carried away trying to show multiple different workflows… just focus on the final worklow you chose, and describe it well enough so your reader understands what you did.

approach/methodology:

Data: you give peeks at your data processing steps without ever explain your exact steps and decisions for data preprocessing. you mention “I also employ a lemmatizer instead of a stemmer after tokenization, but that is unlikely to have a large effect” but no details provide about what your options are or why you chose them.

“I have already tried word2vec and Tfidf vectorizer with confusing results, but I may try again with data organized in a more effective way.” -> vague and non-informative

Very little explained about the models chosen. Rationale seems to be “this other paper chose these models so I will too”. No clear explanation of what hyperparameters you are planning to tune or experimental conditions (which values to try)

Tech Writing:

The first sentence of a paragraph should outline THE ONE idea of the paragraph. don’t tell a chronological story where the reader has no idea what to expect after the first sentence.

Extreme overused of first person. don’t use first person “I” or “My” and avoid first person plural “we” unless necessary to distinguish your work from other work.

“So far, I went down a few rabbit holes in the wrong direction, but I have a clearer path forward. I am taking small steps and trying to keep the Python Notebook tidy and efficient, with helper functions in separate files and processed data stored and read from files instead of re-running steps that have been completed”-> don’t include a diary/journal entry about your experiences. this doesn’t belong in tech writing. think about what you would expect to see in a good published paper, and you quickly realize that this kind of discussion doesn’t belong.

“Below is an example of the dataframe containing…” -> for all figures, tables, etc, use numbered captions and do proper crossrefs (by number) in the main body text when describing them. “Figure 1 displays….”